人工智能辅助下的心理健康新型测评[1]

姜力铭[1] 田雪涛[2] 任 萍[3] 骆 方[1]
([1]北京师范大学心理学部, 北京 100875) ([2]北京交通大学计算机与信息技术学院, 北京 100044)
([3]中国基础教育质量监测协同创新中心, 北京 100875)

摘 要 近年来，人工智能技术的飞速发展及应用催生了“智能化心理健康测评”这一领域。智能化心理健康测评能够弥补传统方法的不足，降低漏诊率并提高诊断效率，这对于心理健康问题的普查及预警具有重大意义。目前，智能化心理健康测评处于初步发展阶段，研究者基于在线行为数据、便携式设备数据等开展主要以数据驱动为导向的探索研究，旨在实现更高的预测准确率，但是测评结果的可解释性等指标尚

够理想。未来的智能化心理健康测评需要强调心理学领域知识和经验的深度介入，提高测评的针对性和精细化程度，加强信效度检验，这对于智能化心理健康测评工具的进一步发展和应用至关重要。

关键词 人工智能，大数据，心理健康，心理测评

1收稿日期: 2021-02-22
* 国家自然科学基金联合基金项目(U1911201)
通信作者: 骆方, E-mail: luof@bnu.edu.cn

1 引言

随着社会的进步和发展加快了人们的生活节奏，也加剧了社会竞争，这些变化必然会对个体的心理健康产生重大影响。在这一时代背景下，如何对心理健康进行高效且精准的测评尤为重要，这是了解民众心理健康状况以及提供有效干预的前提。

近年来，机器学习、深度学习等人工智能和大数据挖掘技术逐渐应用于心理健康领域，带来了心理健康测评方法的革新，也催生了“智能化心理健康测评”这一新兴领域。人工智能是研究并开发用于模拟和延伸人类智能的方法、技术及应用系统的一门科学。机器学习是人工智能最重要的技术手段，旨在探索、建模大量变量之间的复杂高维交互作用(Bzdok & Meyer-Lindenberg, 2018; Kodratoff, 2014)。通过应用人工智能技术进行数据的获取和分析，以及采用机器学习方法表征和建模特征与心理状态之间的关系，智能化心理健康测评能够辅助并一定程度上替代人工测评。与此同时，智能化心理健康测评也大大拓宽了传统心理健康测评手段(如量表法，访谈法等)的测评形式和数据分析方法，使得研究者能够基于更加仿真的任务情境，获取多模态的数据进行协同分析和建模，实现更加高效、精准的测评。因此，本文将针对智能化心理健康测评这一领域的研究进展、目前存在的问题以及未来发展方向进行概述和讨论。

2 智能化心理健康测评的主要研究方向

目前，研究者主要基于社交媒体数据、智能设备数据以及电子游戏数据开展智能化心理健康测评，从大量在线行为数据中挖掘特征或模式，进而实现对心理健康问题的预测。Latynov 和 Shepeleva(2020)提出数字心理测量学(digital psychometrics)这一研究方向，将其界定为根据个体的数字痕迹来预测各种心理特征(如人格特质，情绪状态，价值观，动机等)，基于在线行为数据的心理健康测评可以视为该领域的一类具体任务。

除在线行为数据外，研究者也通过可穿戴设备来采集与心理健康相关的数据，主要包括脑电数据、眼动数据以及运动数据等，通常在实验室中通过专业设备来采集。近年来，研究者尝试基于便携式可穿戴设备采集日常生活中的数据，从中提取特征进行分析和预测。

不同来源的数据具有不同的特点，数据挖掘、分析和建模的方法也存在差异。下面分别对基于社交媒体数据、智能设备数据、电子游戏数据以及可穿戴设备数据开展的智能化心理健康测评研究进行概述，四种数据的简要比较见表 1。

2.1 基于社交媒体数据的心理健康测评

社交媒体数据在心理测评中的应用非常广泛(Kern et al., 2016; Kosinski et al., 2016; Park et al., 2015)。社交媒体上的文本主题开放性高、内容丰富且时间跨度长，承载了个体的思想和情绪情感，对于评估个体心理具有极高的价值(Kern et al., 2016; Mandryk & Birk, 2019)。研究者收集用户在线发布的文本内容，探索有关心理特质以及心理健康状况的表现，采用机器学习和自然语言处理技术构建预测模型。国外研究者主要基于脸书、推特等平台上用户发布的内容来预测其心理特质(Aung & Myint, 2019; Marouf et al., 2019)以及心理健康问题(Eichstaedt et al., 2018)。国内研究者多依托微博、知乎等平台进行相关研究，例如，分析不同生活满意度水平的用户的语言差异(汪静莹 等, 2016)以及检测用户的抑郁、焦虑以及自杀倾向(Cheng et al., 2017)。此外，研究者也尝试利用学生在线学习平台上的写作及评论数据，构建了针对小学生心理特质的预测模型(骆方 等, 2020; 张晗 等, 2020)。

随着人工智能领域相关技术的发展，具有更高性能的深度学习模型不断涌现，提高了智能化测评的准确率(LeCun et al., 2015)。例如，Ive 等人(2018)首次在研究中采用循环神经网络(Recurrent Neural Network, RNN)来预测社交媒体上的帖子中所涉及的心理健康问题，由于 RNN 能够更好地建模具有序列特征的文本数据，其预测结果明显优于以往常用的卷积神经网络(Convolutional Neural Networks, CNN)。然而，模型深度和复杂度的上升往往导致模型可解释性的下降，为解决这一问题，研究者尝试在模型中纳入注意力(attention)机制，自动识别对于预测特定心理健康问题最重要的特征，帮助研究者更好地理解和解释模型结果(Lynn et al., 2020)。可以看出，基于社交媒体数据开展的心理健康测评研究中，研究者始终追求的目标是努力提高模型的预测准确率，但是关于模型的可解释性的问题已经逐渐受到关注。

2.2 基于智能设备数据的心理健康测评

智能手机等便携式电子设备中记录着个体的日常行为数据，包括应用软件的使用、沟通(打电话、发短信)、听音乐、拍照、位置移动(基于 GPS)、连接(蓝牙、WIFI)等，这些行为数据为预测个体心理特质提供了有效的信息。德国慕尼黑大学的研究团队收集了 624

被试连续 30 天的智能手机日志数据，据此构建大五人格的预测模型，识别出了六类对人格特质具有明显预测作用的特征，包括：1)沟通及社交，2)音乐的消费，3)应用的使用，4)位置的移动，5)手机的总体活动，6)日间和夜间活动。该模型的预测结果与效标的相关为 0.4，达到了以往基于社交媒体数据进行人格预测的准确率，显示出基于智能手机日志数据进行心理测评的可行性(Stachl et al., 2020)。

随着智能手机等移动设备的广泛应用，音视频数据的采集和分析也变得更加便利，研究者从中提取动作、语音及表情等特征，实现心理健康的即时、自动评估。音频特征与心理健康状态具有相关性(Cannizzaro, 2004; Mundt, 2012)，研究者尝试基于语音数据筛查心理健康问题。例如，胡斌等人(2018)收集了抑郁症患者以及正常人群在正性、中性以及负性三种情绪状态，以及在语言问答、文本朗读和图片描述三种任务类型下的语音数据，构建了抑郁症的语音识别模型，模型准确率达到 82.9%。Afshan 等人(2018)对抑郁症患者、焦虑症患者以及正常人群的访谈录音进行分析，尝试对心理健康问题进行识别，模型准确率达到 95%。视频中往往记录了个体的面部表情和身体动作，研究者试图通过面部动作编码系统来识别面部肌肉的震颤和变化，捕捉个体的微表情来识别心理健康问题(de Meto et al., 2020; Wang et al., 2018)。Zhao 等人(2019)从视频中提取步态特征来构建预测模型，对情绪的预测准确率达到 80%以上，对焦虑和抑郁的预测结果与效标的相关分别为 0.74 和 0.64。由此可见，随着人工智能技术的发展，心理健康测评将逐渐融入人们的生活中，实现更加便利、高效的评估。

2.3 基于电子游戏数据的心理健康测评

近年来，随着电子游戏的普及，游戏数据也受到了研究者的关注。游戏数据是指玩家在游戏过程中的日志数据，包含了个体在虚拟游戏环境中的丰富的行为表现，研究者可以据此评估个体的能力和心理特质，这类测评方法被称为“基于游戏的测评”(Game-based assessment, GBA) (Heinzen et al., 2015)。基于游戏的测评提供了仿真的交互场景，降低了个体的测验焦虑，同时规避了传统心理测评存在的社会称许性反应等问题，从而获取更为真实的行为表现， (徐俊怡, 李中权, 2021)。

目前，基于游戏的测评主要用于评估个体的能力，例如问题解决能力(Shute et al., 2016)、推理能力(孙鑫 等，2018)、论证推理能力(Song & Sparks, 2019)以及社会情绪能力(DeRosier & Thomas, 2018)等，在认知障碍诊断中也有较多应用(Flynn et al., 2019; Hautala et al., 2020; Manera et al., 2015; Song et al., 2020)。此外，研究者也尝试基于游戏测评某些积极人格，比如依从性(van Nimwegen et al., 2011)和坚持性(DiCerbo, 2014; Ventura & Shute, 2013)等。目前针对心理健康的游戏化测评还非常少，但相关研究正不断涌现，例如 Johannes Dechant 等人(2021) 尝试基于游戏测量个体的社交焦虑水平。

已有基于游戏的测评多数采用商业化游戏数据。基于商业化游戏数据提取的行为、认知和情感等特征可以作为预测个体心理健康的依据(Mandryk & Birk, 2019)。然而，商业化游戏中的娱乐性因素繁多，难以准确地诱发并捕捉特定心理健康问题的行为表现，因而测评结果的可靠性和精细度不足。为实现真正有效的心理健康测评，研究者需要针对研究目

的独立设计游戏或对商业化游戏进行改编，设置能够诱发特定行为的场景和任务，并对相关的行为特征进行埋点记录。

2.4 基于可穿戴设备数据的心理健康测评

心理健康问题往往伴随着明显的生理反应，研究者通过可穿戴设备采集脑电、眼动、心率、皮肤温度等生理指标进行心理健康监测。脑电记录了大脑皮层的电活动，反映了个体对特定刺激的情绪变化(Alhagry et al., 2017; Song et al.,2018)，因此有研究者使用脑电数据来识别与情绪相关的心理健康问题。例如，Deng 等人(2019)采集高情绪障碍者和低情绪障碍者在观看不同情感类型的影片过程中的脑电数据，采用支持向量机构建预测模型，其准确率达到 95.20%。Ay 等人(2019)基于脑电数据构建长短时记忆网络模型(Long Short-Term Memory, LSTM)识别抑郁症患者，模型在左右半球的准确率分别为 97.66%和 99.12%。此外，研究者也基于脑电数据分析个体的注意及认知模式，进而检测与注意相关的心理障碍。例如，Dubreuil-Vall 等人(2020)采用 Flanker 任务收集 ADHD 患者和正常被试的事件相关电位，构建卷积神经网络作为预测模型，模型准确率为 88%±1.12%(Dubreuil-Vall et al., 2020)。除此之外，脑电数据也被用来诊断创伤后应激障碍(Laxminarayan et al., 2020; Meyer et al., 2018)和自闭症(Bosl et al., 2018; Brihadiswaran et al., 2019)等诸多心理健康问题。

通过眼动追踪技术获得的眼动数据也是智能化心理健康测评的一类重要数据。研究者采集被试在特定任务中或刺激下的眼动数据，采用机器学习方法提取凝视时间、凝视移动和瞳孔大小等特征并构建预测模型。例如，De Silva 等人(2019)采集被试在不同事件下的眼动数据，采用决策树算法构建预测模型并实现了 84%的准确率；Zhang 等人(2020)结合脑电数据与眼动数据来识别焦虑症患者，采用支持向量机算法构建预测模型并实现了 82.70%的准确率。清华大学的马惠敏等人基于眼动数据提取被试的注意偏向特征来预测抑郁及焦虑状态，预测模型的准确率、灵敏性和特异性均在 0.8 以上(Pan et al., 2019)。

心理健康与情绪和压力状态具有紧密联系，皮肤温度以及心率等生理指标反应了个体的情绪和压力状态，因而具有预测个体心理健康水平的潜力。例如，采用红外热成像技术测量皮肤温度来检测情绪(Cardone & Merla, 2017)，通过心率和心率变异性等指标检测压力状态(Castaldo et al., 2019; Pereira et al., 2017; Pluntke et al., 2019)和焦虑水平(Ihmig et al., 2020; Wen et al., 2018)。然而，影响个体生理指标的因素众多，生理指标的变化并不完全由心理健康因素造成，研究者需要结合更多监测指标对个体的心理健康状况进行综合判断。

近年来可穿戴设备不断升级，EEG 耳机等小巧的便携式可穿戴设备不断涌现，为个体心理健康状况的持续、无侵扰监测提供了可能(Lo et al., 2017; Richer et al., 2018)。除了利用现有的可穿戴设备外，研究者也尝试针对特定研究问题和目标群体开发专门的可穿戴设备，

例如，中国科学院计算所的陈益强等人与安定医院合作开发了针对儿童注意力缺陷多动症的可穿戴式辅助诊断评估系统，该系统能够感知儿童的敏捷性和冲动性，预测准确率、灵敏性和特异性均达到 0.9 以上(Jiang et al., 2020)。可以看出，研究者们致力于采用更加高效、无侵扰的数据采集方式，实现生态化的、可融入应用场景的心理健康测评并且已经取得了一定进展。

表 1 智能化心理健康测评的四类数据的比较

数据来源	数据获取方式	数据类型	数据量	数据与心理健康研究的相关性	数据在心理健康问题预测中的应用情况
社交媒体	直接爬取公开的社交媒体平台	文本、图像、行为(如点赞、浏览)及元数据(如性别、年龄、位置)等	巨大	不直接相关	有一定的应用，如预测焦虑、抑郁等，预测准确性较低
	在社交媒体上发布相关写作任务，招募被试完成并获取数据		有限	高相关性	
智能设备	招募被试提供数据	通话、短信、听音乐、拍照、位置移动、蓝牙连接、应用软件的使用、音频及视频等	有限	不直接相关	有一定的应用，如预测焦虑、抑郁、自杀倾向等，预测准确性较高
电子游戏	从商业游戏后台直接导出数据	游戏中的行为、发言内容、与其他玩家的互动等	巨大	不直接相关	直接应用非常少，如预测社交焦虑等，但有一些对心理健康相关的心理特质的预测，预测准确性较高
	基于特定研究问题开发或改编游戏，招募被试完成并获取数据		有限	高相关性	
可穿戴设备	招募被试佩戴可穿戴设备，在实验室中完成相关任务，获取数据	脑电、眼动、心率、皮肤温度等生理数据以及精细运动数据	有限	高相关性	应用广泛，如预测焦虑、抑郁、创伤后应激障碍、注意缺陷等，预测准确性高
	招募被试在日常生活中佩戴便携式可穿戴设备，采集日常数据			不直接相关	

3 智能化心理健康测评存在的问题及未来研究方向

智能化心理健康测评是一个新兴的交叉研究领域，目前正处于起步和探索的阶段。该领域的相关研究多数由人工智能及计算机领域的专家主导开展，研究往往基于公开的大规模在线日志数据进行挖掘，从中捕捉与心理健康问题相关的特征及模式并实现预测(Chen & Wojcik, 2016; Kern et al., 2016)。这类研究通常缺乏特定的研究假设，目标是实现更高的模型预测准确率，经常采用数据驱动的研究方法来建模，这就导致预测模型成为一个"黑匣子"，难以为外部行为特征与心理健康的关系提供清晰和明确的解释(Voosen, 2017)。此外，已有研究仅能对个体是否存在某种心理健康问题做二分判断，无法提供细化的评估结果和详细的诊断信息，难以为临床诊断和治疗提供参考。因此，智能化心理健康测评的研究需要强调心理学领域的知识和经验的深度介入，进一步提高测评的针对性、可解释性和精细化水平，加强对测评工具的信效度检验，这对于智能化心理健康测评工具的进一步发展和应用至关重要。

得益于计算机技术的发展，越来越多的机器学习及深度学习算法被封装为程序包，便于心理学研究者直接调用并独立开展心理健康测评的研究(Chen & Wojcik, 2016; Kosinski et al., 2016)。然而，机器学习模型的表现受到诸多环境因素的影响，在实际应用中需要研究者对模型参数进行精细调整甚至针对具体任务开发新的算法模型。因此，智能化心理健康测评系统的搭建需要机器行为(Machine behavior)领域的知识及经验的参与，关注并探究算法在不同条件下的表现(Rahwan et al., 2019)，尽管这并非心理健康测评直接关注的问题，但能够帮助研究者更好地理解和应用人工智能技术，规避预测偏差从而提升测评的有效性。

智能化心理健康测评需要计算机领域与心理学领域的深度融合。一方面，强调心理学领域的知识经验以提高测评的针对性、可解释性和精细化水平，加强对新型测评工具的信效度检验；另一方面，在保证测评的有效性和可靠性的基础上，采用计算机领域的新方法和新进展，获取多模态数据进行协同分析和建模，进一步提升预测准确率。最后，智能化心理健康测评领域的研究者也必须面对隐私和伦理问题。下面就前述主要问题及未来发展方向逐一进行论述。

3.1 强调测评的针对性和精细化

为实现真正高效精准的智能化心理健康测评，研究者需要开展更具有针对性和精细化的研究。对在线行为数据的探索性分析提供了具有启发意义的信息，研究者需要在此基础上定位具体的研究问题，基于理论来设计任务以获取与目标问题高度相关的数据。例如，何祺玮等人(2012; 2017)基于个体的语言表达来识别创伤后应激障碍患者，研究者在心理健康论坛中设置与创伤后应激障碍相关的写作任务，获取被试的自述文本。相比于从社交媒

体获取一般性的文本，针对性的主题写作任务能够更好地激发与 PTSD 相关的文本特征，例如，具有多种创伤后应激障碍的患者文本中包含更多与事件(如“火灾”)以及时间(如“年”)相关的表达，而具有单一创伤后应激障碍患者的文本中包含更多与症状(如“噩梦”)相关的表达，研究者基于文本特征构建的预测模型达到了 80%以上的准确率。

基于可穿戴设备开展的研究大多基于心理学的实验范式进行任务设计，因而研究的针对性通常较高。例如，陈益强等人开发的儿童 ADHD 可穿戴式辅助诊断评估系统基于心理学的 ADHD 实验范式，开发出三大类任务：1)实物交互场景，如手指戳洞任务等；2)屏交互场景，如多目标追踪任务等；3)肢体交互场景，如小鸟喂水任务等。任务覆盖 DSM-5 对 ADHD 的 18 项描述(Jiang et al., 2020)，提取的指标涵盖 ADHD 的各个维度。再如，马惠敏等人(Pan et al., 2019)通过眼动数据预测抑郁及焦虑的研究中采用以反应时为核心的启动、竞争的实验范式，该研究基于明尼苏达多项人格量表(MMPI)以及心理学语义与图像间的映射关系构建了心理图像库，以此作为心理特征提取与分析的素材。该研究不仅能够提供个体心理健康问题的预测结果，也能够输出被试转移时间最长的图像以便研究者进行深入挖掘和根因分析。

研究的精细化包括预测过程的精细化和预测结果的精细化。预测过程的精细化强调有意义的特征提取。目前研究者对于心理健康问题的行为指标已经有了较为明确的认识，但对其脑特征、生理特征及文本特征等还不够了解，深入研究心理健康问题的多元指标将扩展研究者对于目标构念的情感、认知和行为表现的理解(Kern et al., 2016)。预测结果的精细化是指从粗糙的二分诊断逐渐细化到连续、分型诊断。以抑郁症为例，多数研究仅能区分重度抑郁患者和正常人群，为了能够识别轻度抑郁患者并避免其发展为重度抑郁，研究者需要对症状的严重程度进行精细化诊断。北京师范大学的邬霞等人采用 Stroop 任务研究抑郁症患者脑电的功能连通性变化，创新性地将 DTW 算法进行改进并引入到脑网络的构建中，实现了精准刻画线性相关与非线性相关同时存在的脑区信号，并通过层次聚类成功分解得到大脑在执行情绪任务的多尺度脑信号特征(Guo et al., 2017)。研究团队结合 EEG 和 PPG 两种生理信号，综合考虑来自于大脑和外围生理指标中的信息，建立了能够精确评价认知负荷的多生理指标模型(Yu et al., 2018)。该团队还提出了稀疏重叠模块化的高斯图模型算法，不仅能够更准确地估计功能连接网络结构，也明显改善了特征提取的精度，提高了计算机辅助诊断脑疾病的性能(Zhu et al., 2020)。

智能化心理健康测评不仅需要研究者针对特定的心理健康问题，基于心理学理论和范式来设计任务，同时也需要充分利用数据挖掘技术来探索潜在的模式和特征，拓宽对特定心理健康问题的理解。可以看出，智能化心理健康测评要求研究者探索数据驱动与理论驱动相结合的解决方案，这与 von Davier 等人(2013)提出的计算心理测量学的思想不谋而合。计算心理测量学强调基于理论采用自上而下的方式来设计指标，同时引入机器学习方法进行自下而上的数据挖掘(von Davier et al., 2013; 2019)，这一框架目前主要应用于问题解决能

力评估(Polyak et al., 2017)以及学习评估(von Davier et al., 2019)等任务中，在心理健康测评中应用较少。Cipresso 等人(2019)尝试基于计算心理测量学框架检测个体的压力状态，该研究基于领域知识来设定需要获取的生理指标，采用 Stroop 任务和算术任务作为心理压力源，收集被试在静息状态和压力状态下的血容量脉冲、胸腔呼吸和皮肤电导率等生理数据，通过重复方差分析等统计方法检验指标的有效性，最后采用机器学习模型进行预测(Cipresso et al., 2019)。该研究显示出将计算心理测量学应用于心理健康测评中的潜力。尽管目前的智能化心理健康测评的相关研究中很少涉及对计算心理测量学的直接探讨，但一些具有针对性和精细化的研究中已经体现出了计算心理测量学的思想。随着智能化心理健康测评的发展，计算心理测量学应当得到更多的关注和应用。

3.2 引入测量学的证据中心设计

为实现测评的针对性和精细化，研究者需要有针对性地创设任务和情境来激发被试的相关行为指标，获取更加真实、丰富的行为数据。近年来，研究者尝试采用虚拟仿真以及人机交互技术来呈现测验任务，这种测评形式被称为“基于仿真的测评”(Simulation-Based Assessment)(Mislevy, 2013)。与高度结构化的传统测验不同，基于仿真的测评为被试提供了自由探索的环境，收集被试在面对刺激和解决任务时的自发反应，在降低被试的测试焦虑的同时获取更加真实的行为指标。被试在虚拟环境中产生的大量过程性数据也为动态、持续的测评提供了可能(Shute et al., 2016)。

然而，基于仿真的测评在提升测评真实性和生态性的同时也带来了更高的测量误差。过程性数据中混杂着大量与测评目标无关的信息，如果采用无规则的数据挖掘则难以保证测评的有效性，指标提取与测评结果之间的关系也缺乏可解释性。为了在新型测评环境中进行有效测评，Mislevy 等人(2003)提出证据中心设计(Evidence-Centered Design, ECD)。证据中心设计是一种围绕证据的评估设计和评估实施方法，通过任务设计来收集与心理构念相关的证据。证据中心设计包括学生模型、证据模型与任务模型三部分。学生模型回答“测什么”的问题，即依据相关理论定义目标特质的结构。学生模型通常是多维的，包括能力、特质或态度等多个方面(Shute et al., 2011)。证据模型回答“如何测”的问题，确定反映目标特质的指标及计分规则，例如，是否解决了问题、是否使用了特定的工具等。研究者需要基于相关研究基础及知识经验，将证据模型与学生模型进行链接。任务模型解决“用什么测”的问题，在学生模型与证据模型的基础上设计情境、任务形式以及被试的反应方式。任务可以采用多项选择题等简单的形式，也可以采取更复杂、交互性更强的形式。

证据中心设计适用于游戏测评等基于虚拟环境或人机交互的测评任务开发(Shute et al., 2011)，并且已经得到广泛应用(Lee & Recker, 2017; Johannes Dechant et al., 2021; Mislevy & Haertel, 2006; Snow et al., 2019)。智能化心理健康测评的任务设计也应基于证据中心设计，

在学生模型中细化特定心理健康问题的不同维度和分型，提高测评系统的精细化水平；在任务模型中基于特定心理健康问题的典型行为表现来确定指标和计分规则，提高特征提取的有效性和可解释性；在证据模型中参考心理学范式设置测评情境和任务，同时结合虚拟仿真的测评形式，更好地激发被试的相关行为指标。由此可见，证据中心设计的应用将进一步提升智能化心理健康测评的针对性和精细化程度。

3.3 注重测评结果的信效度检验

智能化心理健康测评作为一种新的心理测量方法，需要通过信效度检验以保证测评结果的有效性和科学性。信效度检验回答了预测模型是否测量了目标特质、测量结果是否稳定等一系列重要问题。只有进行了充分的信效度检验，智能化心理健康测评工具才能够得到大规模的应用，尤其是在高利害场景中(如，选拔、考试等)应用以避免较大的争议。

目前，智能化心理健康测评主要采用计算机领域的评估指标，如准确率、召回率等，考虑信效度检验的相关研究非常少见(Tay et al., 2020)。Park 等人(2015)在基于社交媒体数据预测大五人格的研究中检验了重测信度，研究者以 6 个月为单位划分数据，各维度预测结果在相邻两个时间单位间的相关达到 0.70 以上。由于个体的在线行为容易受到网络环境中的诸多因素的影响，因此检验工具的跨时间稳定性十分必要，在未来的相关研究中应尽可能包含这方面的检验结果。

智能化心理健康测评中，机器学习模型充当了评分员的角色，因此，模型选择和构建的恰当性是影响预测结果的重要因素。Sajjadiani 等人(2019)根据传统的评分者一致性信度(inter-rater reliability)提出了算法一致性信度(inter-algorithm reliability)，检验不同模型在同一批数据上的评分一致性。由于每种模型都存在优势和弊端，研究者应结合具体任务进行模型选择并对适当的备选模型进行检验和比较。

机器学习模型容易对单次获取的训练集数据过分拟合，因此智能化心理健康测评需要考虑预测模型的泛化性能。研究者通常采用交叉验证方法对模型的泛化能力和稳定性进行估计(Kosinski et al., 2016)。交叉验证方法将样本数据随机分为 K 个大小相似的组，每次以其中一组用作测试集，其它 K-1 组作为训练集，以 K 次测试结果的平均值作为模型准确率的估计。此外，研究者也需要验证工具在不同情境中的泛化能力和普适性。不同的社交媒体平台具有不同的特点，例如，推特主要服务于大众信息的传播而脸书主要服务于熟人之间的交流，这些特点均对个体特质的表现产生影响(Saef et al., 2018)，研究者应采用其它样本数据或研究设计来验证原有发现(Kern et al., 2016)。例如，中国科学院计算所的朱廷劭等人检验了抑郁症患者与正常人群的语音差异的跨情境稳定性，研究表明抑郁者和正常人群之间的语音差异在不同情境下普遍存在，并且识别出差异最大的 12 个重要特征(Wang et.al., 2019)。因此，智能化心理健康测评应重点捕捉具有跨情境稳定性的普遍特征，同时考虑虚拟环境对个体行为表现的影响以提升测评的有效性和可解释性。

目前，智能化心理健康测评只能做到粗筛，无法直接用于诊断，但加强测评的精细化和针对性将有助于提高评估的准确率，同时提供更加丰富的信息帮助医生进行进一步的临床评估和诊断。

3.4 融合多模态数据进行协同分析

随着大数据时代的到来，数据的共享为多模态数据的整合分析提供了可能。个体的心理健康状况通过语言、肢体动作、面部表情、生理反应等多种途径表现，不同的数据来源和

标有其独特优势，综合分析各类数据将实现更加全面和稳健的评估。计算机领域的多模态数据分析方法为智能化心理健康测评带来了革新，研究者开始尝试更多元的数据采集形式，获取多模态的数据进行融合建模，从而发挥信息的互补作用。例如，Williamson 等人(2016)的研究中融合了生理、语音、面孔以及语义四类特征构建抑郁症的预测模型；斯坦福大学的李飞飞等人(2018)利用面部表情以及语音数据构建抑郁症的预测模型；华中科技大学的陈敏等人采集多场景(工作、学习、娱乐)下的多模态数据(脑电、视频、眼动)，构建多动症儿童的注意力评估模型(Chen et al., 2019)。上述研究结果显示，包含多模态数据的模型往往实现了最优的预测效果。

游戏能够同时记录玩家的行为、认知、运动、社交以及情感等多种心理健康

，基于游戏的测评有望成为多模态数据的重要应用场景。已有研究者通过分析游戏中的发言内容来预测玩家的心理健康(Mandryk & Birk, 2019)，通过游戏手柄中的传感器获取生理数据来分析玩家的情绪和认知状态(Mandryk et al., 2013)，通过玩家在游戏中敲击按钮的压力大小来推断其心理健康(Vogel, 2018)等。不同类别的数据和指标反映了心理健康的不同侧面，全面收集各类生理、心理及行为数据进行协同建模和综合判断，这对于心理健康问题的精准筛查至关重要。

3.5 对隐私及伦理等问题的考虑

目前，国内外关于智能化心理健康测评的研究尚处于初步阶段，随着人工智能与大数据技术的发展，相关研究的伦理问题将逐步受到重视。基于在线行为数据的研究中，被试往往无法得知自己的信息已被用于研究，未来研究中的数据获取和使用应尽可能使被试知情。此外，智能化心理健康测评必须考虑被试的隐私保护，规避隐私信息泄露的风险。传统测评中研究者能够通过删除被试的身份信息来保护被试隐私，然而在线行为数据中包含的个人信息难以完全剔除(Kern et al., 2016)。随着研究获取的数据来源的扩展以及信息之间的融合，个体身份的识别将更加容易(Berman, 2013)。研究者应站在被试的角度上考虑哪些

数据可以获取和分析，仅采集研究必需的信息(Kern et al., 2016)，例如，Markus 等人(2020)通过个体的语音数据来评估心理状态的研究中，研究者仅获取语音数据的参数而无法得到原始的语音内容，这样的数据采集及处理方式值得借鉴。

4 结语

心理健康问题的智能化测评是人工智能领域与心理学及医学领域的交叉问题，跨学科的深度交流和共同努力至关重要(Kern et al., 2016)。领域间的深度融合和思维碰撞能够激发出更多的研究成果，惠及人类的心理健康和幸福生活。本文所介绍的研究绝大多数是智能化心理健康测评领域的初步探索，相关研究成果为未来研究提供了基线标准，研究者构建的数据库也为未来研究提供了进一步探索的宝贵资源。

近年来，智能化心理健康测评的研究问题从最常见的抑郁症、焦虑症，扩展到注意力缺陷多动症、创伤性应激障碍、自闭症等各类心理健康问题。如今，AI 不仅仅能够增强人类的能力，使人们看到更多、听到更多，帮助人类思考和计算，同时，AI 也逐渐变得更加有温度，更加关注人类的情绪与情感、人类的心理健康及主观幸福感等。相信未来的心理测评在变得更加智能化的同时也必然变得更加人性化。

参考文献

骆方, 姜力铭, 田雪涛, 肖梦格, 马彦珍, 张生. (2021). 小学生羞怯特质预测及语言风格模型构建. *心理学报*, *53*(2), 155–169.

潘玮, 汪静莹, 刘天俐, 刘晓倩, 刘明明, 胡斌, 朱廷劭. (2018). 基于语音的抑郁症识别. *科学通报*, *63*(20), 2081–2092.

孙鑫, 黎坚, 符植煜. (2018). 利用游戏 log-file 预测学生推理能力和数学成绩——机器学习的应用. *心理学报*, *50*(7), 761–770.

汪静莹, 朱廷劭, 郝碧波, 刘天俐. (2016). 微博用户生活满意度微博语言及行为特征分析. *中国公共卫生*, *32*(2), 225–229.

徐俊怡, 李中权. (2021). 基于游戏的心理测评. *心理科学进展*, *29*(3), 394–403.

张晗, 贾甜远, 骆方, 张生, 邬霞. (2020). 面向网络文本的 BERT 心理特质预测研究. *计算机科学与探索*, *0*, 1–13.

Afshan, A., Guo, J., Park, S. J., Ravi, V., & Alwan, A. (2018). Effectiveness of voice quality features in detecting depression. *Interspeech*, 1676–1680.

Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *Emotion*, *8*(10), 355–358.

Aung Z. M. M., & Myint P. H. (2019 July). Personality prediction based on content of facebook users: A literature review. *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (pp. 34–38) IEEE.

Ay, B., Yildirim, O., Talo, M., Baloglu, U. B., Aydin, G., Puthankattil, S. D., & Acharya, U. R. (2019). Automated depression detection using deep representation and sequence learning with EEG signals. *Journal of Medical Systems*, *43*(7), 1–12.

Berman, J. J. (2013). *Principles of big data: Preparing, sharing, and analyzing complex information*. Waltham, MA: Elsevier.

Bosl, W. J., Tager-Flusberg, H., & Nelson, C. A. (2018). EEG analytics for early detection of autism spectrum disorder: A data-driven approach. *Scientific Reports*, *8*(1), 1–20.

Brihadiswaran, G., Haputhanthri, D., Gunathilaka, S., Meedeniya, D., & Jayarathna, S. (2019). EEG-based processing and classification methodologies for autism spectrum disorder: A review. *Journal of Computer Science*, *15*(8).

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 223–230.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, *56*(1), 30–35.

Cardone, D., & Merla, A. (2017). New frontiers for applications of thermal infrared imaging devices: Computational psychopshysiology in the neurosciences. *Sensors*, *17*(5), 1042.

Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*, *19*(1), 1–13.

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, *21*(4), 458–474.

Chen M., Cao Y., Wang R., Li, Y., Wu, D., & Liu, Z. (2019). DeepFocus: Deep encoding brainwaves and emotions with multi-scenario behavior analytics for human attention enhancement. *IEEE Network*, *33*(6), 70–77.

Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study. *Journal of Medical Internet Research*, *19*(7), e243.

Cipresso, P., Colombo, D., & Riva, G. (2019). Computational psychometrics using psychophysiological measures for the assessment of acute mental stress. *Sensors*, *19*(4), 781.

Deng, Y., Wu, F., Du, L., Zhou, R., & Cao, L. (2019, March). EEG-based identification of latent emotional disorder using the machine learning approach. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 2642–2648). IEEE.

de Meto, W. C., Granger, E., & Lopez, M. B. (2020, May). Encoding temporal information for automatic depression recognition from facial analysis. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1080–1084). IEEE.

DeRosier, M. E., & Thomas, J. M. (2018). Establishing the criterion validity of Zoo U's game-based social emotional skills assessment for school-based outcomes. *Journal of Applied Developmental Psychology*, *55*, 52–61.

De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., Michalek, A. M., & Jayawardena, G. (2019, July). A rule-based system for ADHD identification using eye movement data. *2019 Moratuwa Engineering Research Conference (MERCon)* (pp. 538–543). IEEE.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Educational Technology and Society*, *17*(1), 17–28.

Dubreuil-Vall, L., Ruffini, G., & Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Frontiers in Neuroscience, 14*, 251.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., ... & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, *115*(44), 11203–11208.

Flynn, R. M., Colón-Acosta, N., Zhou, J., & Bower, J. (2019). A game-based repeated assessment for cognitive monitoring: Initial usability and adherence study in a summer camp setting. *Journal of Autism and Developmental Disorders*, *49*(5), 2003–2014.

Guo, Z., Wu, X., Liu, J., Yao, L., & Hu, B. (2018). Altered electroencephalography functional connectivity in depression during the emotional face-word Stroop task. *Journal of neural engineering*, *15*(5), 056014.

Haque, A., Guo, M., Miner, A. S., & Fei-Fei, L. (2018). Measuring depression symptom severity from spoken language and 3D facial expressions. *arXiv preprint arXiv:1811.08592*.

Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J. M., & Richardson, U. (2020). Identification of reading difficulties by a digital game-based assessment technology. *Journal of Educational Computing Research*, *58*(5), 1003–1028.

He, Q., Veldkamp, B. P., Glas, C. A., & de Vries, T. (2017). Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. *Assessment*, *24*(2), 157–172.

He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self narratives: A text mining approach. *Psychiatry Research*, *198*(3), 441–447.

Heinzen, T. E., Landrum, R. E., Gurung, R. A., & Dunn, D. S. (2015). Game-based assessment: The mash-up we've been waiting for. *Gamification in Education and Business* (pp. 201–217). Springer.

Ihmig, F. R., Neurohr-Parakenings, F., Schäfer, S. K., Lass-Hennemann, J., & Michael, T. (2020). On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals. *Plos One*, *15*(6), e0231517.

Ive, J., Gkotsis, G., Dutta, R., Stewart, R., & Velupillai, S. (2018, June). Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 69–77).

Jiang, X., Chen, Y., Huang, W., Zhang, T., Gao, C., Xing, Y., & Zheng, Y. (2020, April). WeDA: Designing and evaluating a scale-driven wearable diagnostic assessment system for children with ADHD. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Johannes Dechant, M., Frommel, J., & Mandryk, R. (2021, May). Assessing social anxiety through digital biomarkers embedded in a gaming task. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, *21*(4), 507–525.

Kodratoff, Y. (2014). *Introduction to machine learning*. San Mateo, CA: Morgan Kaufmann.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, *21*(4), 493–506.

Latynov, V., & Shepeleva, E. (2020). Applied aspects of the use of algorithms of digital psychometrics. *Psikhologicheskii Zhurnal*, *41*(4), 66–77.

Laxminarayan, S., Wang, C., Oyama, T., Cashmere, J. D., Germain, A., & Reifman, J. (2020). Identification of veterans with PTSD based on EEG features collected during sleep. *Frontiers in Psychiatry*, *11*.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, J. E., & Recker, M. (2017). Measuring students' use of self-regulated learning strategies from learning management system data: An evidence-centered design approach. *Analytics for Learning*, 1–14.

Lo, J. C., Sehic, E., & Meijer, S. A. (2017). Measuring mental workload with low-cost and wearable sensors: Insights into the accuracy, obtrusiveness, and research usability of three instruments. *Journal of Cognitive Engineering and Decision Making*, *11*(4), 323–336.

Lynn, V., Balasubramanian, N., & Schwartz, H. A. (2020, July). Hierarchical modeling for user personality prediction: The role of message-level attention. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5306–5316).

Mandryk, R. L., Dielschneider, S., Kalyn, M. R., Bertram, C. P., Gaetz, M., Doucette, A., ... & Keiver, K. (2013, June). Games as neurofeedback training for children with FASD. *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 165–172).

Mandryk, R. L., & Birk, M. V. (2019). The potential of game-based digital biomarkers for modeling mental health. *Journal of Medical Internet Research*, *6*(4), e13485.

Manera, V., Petit, P. D., Derreumaux, A., Orvieto, I., Romagnoli, M., Lyttle, G., ... Robert, P. H. (2015). "Kitchen and cooking", a serious game for mild cognitive impairment and Alzheimer's disease: A pilot study. *Frontiers in Aging Neuroscience*, 7–24.

Marouf, A. A., Hasan, M. K., Mahmud, H. (2019 February). Identifying neuroticism from user generated content of social media based on psycholinguistic cues. *2019 International Conference on Electrical, Computer and Communication Engineering* (pp. 1–5).

Meyer, T., Quaedflieg, C. W., Weijland, K., Schruers, K., Merckelbach, H., & Smeets, T. (2018). Frontal EEG asymmetry during symptom provocation predicts subjective responses to intrusions in survivors with and without PTSD. *Psychophysiology*, *55*(1), e12779.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1).

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.

Mislevy, R. J. (2013). Evidence-centered design for simulation-based assessment. *Military Medicine*, *178*(10), 107–114.

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587.

Pan, Z., Ma, H., Zhang, L., & Wang, Y. (2019). Depression detection based on reaction time and eye movement. *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952.

Pereira, T., Almeida, P. R., Cunha, J. P., & Aguiar, A. (2017). Heart rate variability metrics for fine-grained stress level assessment. *Computer Methods and Programs in Biomedicine*, *148*, 71–80.

Pluntke, U., Gerke, S., Sridhar, A., Weiss, J., & Michel, B. (2019, July). Evaluation and classification of physical and psychological stress in firefighters using heart rate variability. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2207–2212). IEEE.

Polyak, S. T., von Davier, A. A., & Peterschmidt, K. (2017). Computational psychometrics for the measurement of collaborative problem-solving skills. *Frontiers in Psychology*, *8*, 2029.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

Richer, R., Zhao, N., Amores, J., Eskofier, B. M., & Paradiso, J. A. (2018, July). Real-time mental state recognition using a wearable EEG. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5495–5498). IEEE.

Saef, R., Woo, S. E., Carpenter, J., & Tay, L. (2018). Fostering socio-informational behaviors online: The interactive effect of openness to experience and extraversion. *Personality and Individual Differences*, *122*, 93–98.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *The Journal of Applied Psychology*, *104*, 1207–1225.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*, *55*(2), 503–524.

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117.

Snow, E., Rutstein, D., Basu, S., Bienkowski, M., & Everson, H. T. (2019). Leveraging evidence-centered design to develop assessments of computational thinking practices. *International Journal of Testing*, *19*(2), 103–127.

Song, H., Yi, D. J., & Park, H. J. (2020). Validation of a mobile game-based assessment of cognitive control among children and adolescents. *Plos One*, *15*(3), 1–18.

Song, T., Zheng, W., Song, P., & Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *11*(3), 532–541.

Song, Y., & Sparks, J. R. (2019). Measuring argumentation skills through a game-enhanced scenario-based assessment. *Journal of Educational Computing Research*, *56*(8), 1324–1344.

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., ... & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*(30), 17680–17687.

Tay, L., Woo, S. E., Hickman, L., & Saef, R. M. (2020). Psychometric and validity issues in machine learning approaches to personality assessment: A focus on social media text mining. *European Journal of Personality*, *34*(5), 826–844.

van Nimwegen, C., van Oostendorp, H., Modderman, J., & Bas, M. (2011). A test case for GameDNA: Conceptualizing a serious game to measure personality traits. *16th International Conference on Computer Games (CGAMES)* (pp. 217–222). IEEE.

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, *29*(6), 2568–2572.

Vogel, & Lauren. (2018). AI opens new frontier for suicide prevention. *Canadian Medical Association Journal*, *190*(4), E119–E119.

von Davier, A. A., & Halpin, P. F. (2013). Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series*, *2013*(2), i–36.

von Davier, A. A., Deonovic, B., Yudelson, M., Polyak, S. T., & Woo, A. (2019, July). Computational psychometrics approach to holistic learning and assessment systems. *Frontiers in Education*, *4*, 69

Voosen, P. (2017). The AI detectives. *Science*, *357*, 22–27.

Wang, Q., Yang, H., & Yu, Y. (2018). Facial expression video analysis for depression detection in Chinese patients. *Journal of Visual Communication and Image Representation*, *57*, 228–233.

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: A cross-situation study. *BMC Psychiatry*, *19*(1), 1–12.

Wen, W., Liu, G., Mao, Z. H., Huang, W., Zhang, X., Hu, H., ... & Jia, W. (2018). Toward constructing a real-time social anxiety evaluation system: Exploring effective heart rate features. *IEEE Transactions on Affective Computing*, *11*(1), 100–110.

Williamson, J. R., Godoy, E., Cha, M., Schwarzentruber, A., Khorrami, P., Gwon, Y., ... & Quatieri, T. F. (2016, October). Detecting depression using vocal, facial and semantic communication cues. *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge* (pp. 11–18).

Yu, H., Li, F., Wu, T., Li, R., Yao, L., Wang, C., & Wu, X. (2018). Functional brain abnormalities in major depressive disorder using the Hilbert-Huang transform. *Brain Imaging and Behavior*, *12*(6), 1556–1568.

Zhang, X., Pan, J., Shen, J., Din, Z. U., Li, J., Lu, D., ... & Hu, B. (2020). Fusing of electroencephalogram and eye movement with group sparse canonical correlation analysis for anxiety detection. *IEEE Transactions on Affective Computing*, *99*, 1.

Zhao, N., Zhang, Z., Wang, Y., Wang, J., Li, B., Zhu, T., & Xiang, Y. (2019). See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data. *Plos One*, *14*(5), e0216591.

Zhu, Z., Zhen, Z., Wu, X., & Li, S. (2020). Estimating functional connectivity by integration of inherent brain function activity pattern priors. *IEEE/ACM transactions on computational biology and bioinformatics.*

A New Type of Mental Health Assessment Using Artificial Intelligence Technique

JIANG Liming[1], TIAN Xuetao[2], REN Ping[3], LUO Fang[1]
([1] School of Psychology, Beijing Normal University, Beijing 100875, China)
([2] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)
(3 Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing 100875, China)

Abstract: The rapid development and application of artificial intelligence technology has promoted the intelligentization of mental health assessment. Being intelligent could solve the issues of traditional mental health assessment methods and decrease the rate of misdiagnosis and improve diagnosis efficiency, which is critical to the general investigation and early warning of mental health problems. Currently, an intelligent mental health assessment is in the initial stage of development. Related studies have explored the field mainly driven by data, in which researchers use online behavioral data and data from portable devices, aiming to achieve a higher prediction accuracy. However, the interpretability of assessment results is not yet ideal. In view of these

problems, more emphasis should be laid on the knowledge and experience in the field of psychology, by which the research could be more pertinent, refined, reliable, and valid. These are essential directions for the further development and application of intelligent mental health assessment.
Key words: artificial intelligence, big data, mental health, psychological assessment